



## Contribution to the Symposium: 'Marine Acoustics Symposium'

### Original Article

# Increasing the accessibility of acoustic data through global access and imagery

Carrie C. Wall<sup>1,2\*</sup>, J. Michael Jech<sup>3</sup>, and Susan J. McLean<sup>2</sup>

<sup>1</sup>Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, 216 UCB, Boulder, CO 80309, USA

<sup>2</sup>NOAA National Centers for Environmental Information, 325 Broadway, Boulder, CO 80305, USA

<sup>3</sup>NOAA National Marine Fisheries Service, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, USA

\*Corresponding author: tel: +1 303 497 4459; fax: +1 303 497 6513; e-mail: [carrie.bell@colorado.edu](mailto:carrie.bell@colorado.edu)

Wall, C. C., Jech, J. M., and McLean, S. J. Increasing the accessibility of acoustic data through global access and imagery. – ICES Journal of Marine Science, 73: 2093–2103.

Received 27 August 2015; revised 9 January 2016; accepted 20 January 2016; advance access publication 6 March 2016.

The National Oceanographic and Atmospheric Administration (NOAA) uses water column sonar data to assess physical and biological characteristics from the ocean surface to the seabed. Acoustic surveys produce large volumes of data that can deliver valuable information beyond their original collection purpose if the data are properly managed, discoverable, and accessible to the public. NOAA's National Centers for Environmental Information, in partnership with NOAA's National Marine Fisheries Service and the University of Colorado, have created a national archive for water column sonar data to help achieve these goals. Through these efforts, over 21 TB of sonar data are now publicly available. Raw sonar files are difficult to interpret due to their size, complexity, and proprietary format. In order for users to understand the quality and composition of large volumes of archived data more easily, several visualization products were explored. Three processing methods were applied to multifrequency single-beam data (Simrad EK60) collected off the US northwest coast between 2007 and 2013. One method illustrates these complex data in a single image using a novel colour scale [multifrequency single-beam imaging (MFSBI)], another examines the nautical area scattering coefficients between two frequencies ( $\Delta$ NASC), and the third indices the data into acoustic classifications [multifrequency indicator (MFI)]. The ability to apply the algorithms efficiently to multiyear datasets was explored. MFSBI proved effective at conveying the composition of the data and was easily adaptable to automated processing.  $\Delta$ NASC, which required manual seabed corrections, illustrated a generalized pattern for changes in the water column across the shelf. MFI provided an empirically based statistical approach but will require more effort in the near term to evaluate and assess the accuracy and precision of each classification. Overall, spatio-temporal patterns of the acoustic backscatter identified large interannual variations in composition with the continental shelf break often playing a key role in attracting biological assemblages.

**Keywords:** data visualization, marine acoustics, single-beam echosounders, water column sonar data.

## Introduction

Active acoustic (sonar) technologies are of increasing importance for studies examining aquatic ecosystems. Sonars that focus on the water column, the volume of ocean from the near surface to near the seabed, are used to estimate biomass (Simmonds and MacLennan, 2005), conduct trophic- and species-level identification (Benoit-Bird and Au, 2001; Foote, 2009), measure school and patch morphology and behaviour (Soria *et al.*, 1996; Simmonds and MacLennan, 2005), and characterize habitat for commercially and ecologically important fish and invertebrate species (Hutin *et al.*, 2005; Cutter *et al.*, 2010; Pirtle *et al.*, 2015). In addition, these data can be used for seabed characterization supporting both essential fish habitat and safe

navigation, and to map natural methane seeps and undersea oil plumes (Weber *et al.*, 2012, 2014; Skarke *et al.*, 2014). Single and multibeam echosounders using single frequency, multiple frequencies, broadband, or wideband technology are employed to acquire these data (Misund, 1997; Simmonds and MacLennan, 2005; Lavery *et al.*, 2010; Stanton *et al.*, 2010).

These sonar systems deliver valuable information for ecosystem-based fisheries management, but they also produce large data volumes that are costly and complicated to maintain. In collaboration with the National Oceanographic and Atmospheric Administration (NOAA) National Marine Fisheries Service (NMFS) and the University of Colorado, the NOAA National

Centers for Environmental Information (NCEI) are archiving acoustic data collected from NOAA and academic fleets. Through these efforts, raw water column sonar data exceeding 21 TB are now available to researchers and the public around the world. This volume will continue to grow as more data are collected and submitted to NCEI. The metadata associated with the raw acoustic data ensure proper documentation of how, when, why, and where the data were collected. Standard formatting of the metadata, International Organization for Standardization (ISO) 19115 and International Council for the Exploration of the Sea (ICES) fisheries acoustic metadata (ICES, 2014), provides consistency across datasets and is therefore essential for collaboration across agencies. In addition, these metadata provide the basis for NCEI issuance of digital object identifiers (DOIs) to aid in proper citation to the original principal investigator when data are reused.

NCEI's acoustic archive inclusive of metadata ensures these valuable data are preserved for decades to come and aids NMFS in meeting many data stewardship mandates. Additional benefits of this central repository include global access to an unprecedented dataset, increased potential for cross-institution collaboration, and the ability for researchers to address cross-cutting scientific questions to advance the field of marine ecosystem acoustics. Therefore, the value of the archived data is further increased by the ability to use them for scientific research beyond their original collection purpose.

Water column sonar data archived at NCEI can be discovered, queried, and requested through a data access web page ([http://maps.ngdc.noaa.gov/viewers/water\\_column\\_sonar](http://maps.ngdc.noaa.gov/viewers/water_column_sonar)). While the displayed cruise tracks illustrate the spatial extent where surveys were conducted, the quality and composition of the data cannot be discerned. Further, water column sonar data files are complex and come in a proprietary format that requires specialized acoustic processing software or extensive knowledge on how to read the files using a scientific programming language. Visualization tools that allow users of varying backgrounds to quickly and easily understand the content of these complex data are needed. Such tools will allow users to know whether the data are in fact of interest and relevant to their research objectives before going through the process to request and receive the data, which could amount to many terabytes. These steps save time for both the user and the archive staff, and ultimately lead to a more efficient process.

The overarching goal of this research is to test the concepts for several visualization products that can be implemented into the NCEI water column sonar data access web page. We anticipate such interactive science will increase the value of the archived data. The visualization products' metadata will follow the newly established ICES standard for processed acoustic data (ICES, 2014) furthering their global application. Illustrations of processed water column sonar data have been incorporated into the Integrated Marine Observing System single-frequency single-beam data hosted on the Australian Ocean Data Network (<http://portal.aodn.org.au/aodn/>). Though still in its infancy, the Southern Ocean Network of Acoustics (SONA; <http://www.sona.aq/data/data-access/>) is working towards similar goals of making bioacoustic data easily accessible. Ocean observing systems (OOS) established at the global (GOOS) and regional levels [e.g. Gulf of Mexico Coastal Ocean Observing System (GCOOS)] are key potential venues to integrate meaningful visualizations for scientists, namely non-acousticians, to access and digest the acoustic data while exploring other datasets. Here, we examine three classification techniques applied to multifrequency acoustic data to determine their potential for widespread dissemination and interpretation.

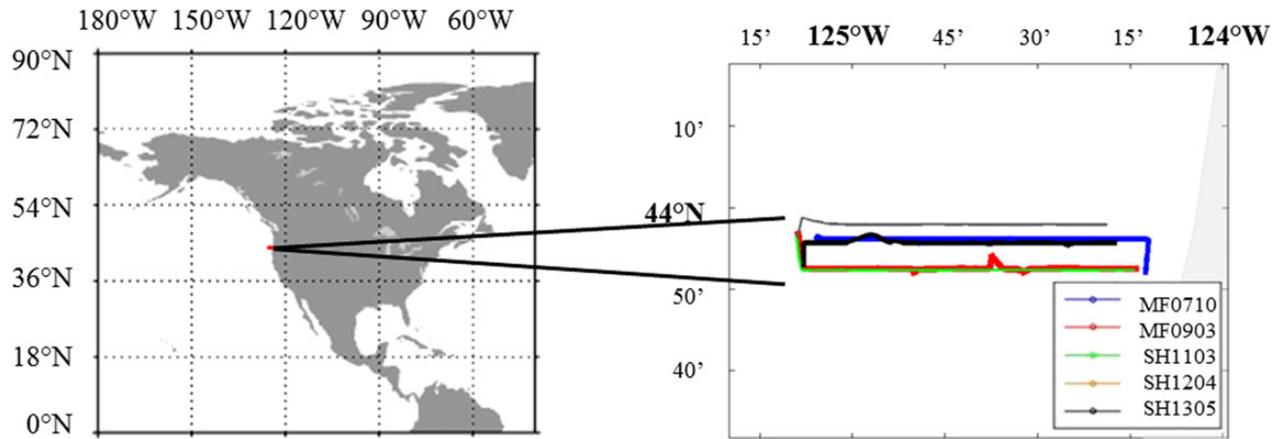
## Methods

Volume backscattering acoustic data ( $S_v$ , dB re  $1 \text{ m}^{-1}$ ) previously collected by the NOAA Northwest Fisheries Science Center (NWFSC) off the northwest coast of the United States were selected to examine data processing concepts across multiple years. The NWFSC conducts acoustic surveys in summer typically every 2 years using systematic transects that result in overlapping cruise tracks between surveys at approximately the same time of year. To gain a better understanding of the content and trophic structures, however ephemeral, along the northwest coast, a subset of Simrad EK60 (Andersen, 2001) acoustic data files collected by the NWFSC from 2007 to 2013 were processed using (i) an adaptation of the Jech and Michaels (2006) algorithm with a novel colour scale applied, termed multifrequency single-beam imaging (MFSBI), (ii) the frequency-difference nautical area scattering coefficient ( $\Delta\text{NASC}$ ), and (iii) the multifrequency indicator (MFI) algorithm adapted from Trenkel and Berger (2013). The results of the three methods (MFSBI,  $\Delta\text{NASC}$ , and MFI) were examined to determine onshore/offshore changes in composition, especially with respect to the continental shelf break, and annual changes within and across categories.

A region of biological interest along the west coast of the US was identified near the 44th parallel ( $43.872^\circ\text{N}$ ,  $125.172^\circ\text{W}$ ;  $43.975^\circ\text{N}$ ,  $124.211^\circ\text{W}$ ; Figure 1). On-transect EK60 files recorded in this region were extracted from the water column sonar data archive's data access web page for NWFSC cruises conducted in 2007 (vessel cruise ID: MF0710), 2009 (MF0903), 2011 (SH1103), 2012 (SH1204), and 2013 (SH1305; NWFSC, 2007, 2009, 2011, 2012, 2013). The NOAA Ship Miller Freeman [hereafter Freeman (MF)] was used in the 2007 and 2009 cruises and the NOAA Ship Bell M. Shimada [hereafter Shimada (SH)] for the 2011, 2012, and 2013 cruises. The Freeman employed four transducers (18, 38, 120, and 200 kHz) compared with the Shimada's five (18, 38, 70, 120, and 200 kHz). To allow for a direct comparison between the two vessels' datasets, the Shimada's 70 kHz data were omitted from these analyses. The period when files were collected within the region of interest for each cruise is outlined in Table 1.

All vessels were calibrated using the standard sphere method (Foote *et al.*, 1987; Simmonds and MacLennan, 2005). Target strength and echo integration data were collected to calculate echosounder gain parameters to ensure the quality of the system performance. On-axis (Simmonds and MacLennan, 2005) and beam-pattern measurements were also taken. To minimize the effect of surface bubbles and transducer "ring down", acoustic data were collected from 14 m below the surface,  $\sim 5$  m below the centreboard-mounted transducer face. Data were collected to a maximum depth of 600 m in 2009 and 750 m for the remaining cruises. The transducer depth was accounted for so that the depth of each  $S_v$  sample is relative to the sea surface. This recording range resulted in a ping rate of  $\sim 1 \text{ ping s}^{-1}$  for 2009 and 1 ping per  $1.1 \text{ s}^{-1}$  for the remaining cruise.

All files were preprocessed by first aligning pings in the time/distance domain across the frequency components. Data were then binned vertically to 1000 data points between 0 and 750 m (i.e.  $S_v$  at 0.75 m intervals). Noise filters were applied to remove background noise and intermittent impulsive noise. Background noise was removed following De Robertis and Higgenbottom (2007) where the signal-to-noise ratio was set to 10 dB. Impulsive noise "spikes" have a short duration of  $< 1$  "ping" or transmit-receive cycle (Ryan *et al.*, 2015), so a two-sided comparison method is applied where  $n$  pings on either side of the current ping are examined. Similar to the methods in Ryan *et al.* (2015), an  $n$  of 1 was applied and the



**Figure 1.** On transect Simrad EK60 files recorded within the region of interest ( $43.872^{\circ}\text{N}$ ,  $125.172^{\circ}\text{W}$ ;  $43.975^{\circ}\text{N}$ ,  $124.211^{\circ}\text{W}$ ) for NWFSC acoustic cruises conducted in 2007 (MF0710), 2009 (MF0903), 2011 (SH1103), 2012 (SH1204), and 2013 (SH1305). Not all cruise tracks can be discerned easily due to the high level of overlap. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.

**Table 1.** NWFSC acoustic survey information.

Cruise ID	Year	Dates	Time (UTC)	No. of files
MF0710	2007	8 July	01:02 – 16:50	10
MF0903	2009	24 July	02:24 – 18:32	8
SH1103	2011	21 – 22 July	21:00 – 00:53	12
SH1204	2012	24 July	19:16 – 23:53	48
SH1305	2013	29 July	16:03 – 23:35	25

Provided are the cruise IDs, time frame, and number of raw acoustic files (No. of files) recorded within the region of interest ( $43.872^{\circ}\text{N}$ ,  $125.172^{\circ}\text{W}$ ;  $43.975^{\circ}\text{N}$ ,  $124.211^{\circ}\text{W}$ ). Cruise IDs beginning with MF denotes the Freeman and SH represents the Shimada.

current ping was removed if the  $S_v$  was 10 dB higher or lower than the adjacent pings. Each excised data point was replaced with the local mean  $S_v$  calculated using the 7  $S_v$  values at the same range (i.e. same row of the echogram). To smooth and reduce stochastic variability in the data, a  $3 \times 3$  convolution filter with the maximum weight on the centre pixel and summed kernel weights equal to one was then applied. Echograms of the processed 38 kHz frequency data for all years are provided in Figure 2.

### Multifrequency single-beam imaging

The MFSBI algorithm enables multiple frequency data to be illustrated in a single image by depicting the dominating frequency or frequencies. A threshold of  $-66$  dB was applied to the preprocessed  $S_v$  data which was empirically established through inspection and served to remove low-amplitude backscatter (Jech and Michaels, 2006). The  $S_v$  echograms were then transformed to a matching array of unique values where  $S_v$  greater than  $-66$  dB were assigned a unique, positive value based on the acoustic frequency while data below the  $S_v$  threshold were set to 0. Pixels above the threshold within the 18 kHz echogram were set to 1, pixels in the 38 kHz echogram were set to 3, 70 kHz were set to 29, 120 kHz were set to 7, and 200 kHz were set to 13. These integers were chosen as the summation of any combination of the numbers will produce a unique result. This is an important aspect as the four matrices of values representing each frequency component were then summed together to create a single matrix.

Applying the colour scale to the data, values above the threshold for 18 kHz were set to light grey, blue for 38 kHz, red for 120 kHz,

and yellow for 200 kHz (Figure 3). The summation of the matrices results in pixels with a unique colour. Using the colour wheel as the basic concept, a pixel in the summed matrix that consisted of both 38 kHz (blue) and 120 kHz (red) will be coloured purple. Similarly, the combination of 120 kHz (red) and 200 kHz (yellow) will result in an orange pixel. The addition of light grey (18 kHz) to any combination will produce a less saturated colour (e.g. a data point in the summed matrix that contains components from 18, 38, and 120 kHz will be coloured light purple).

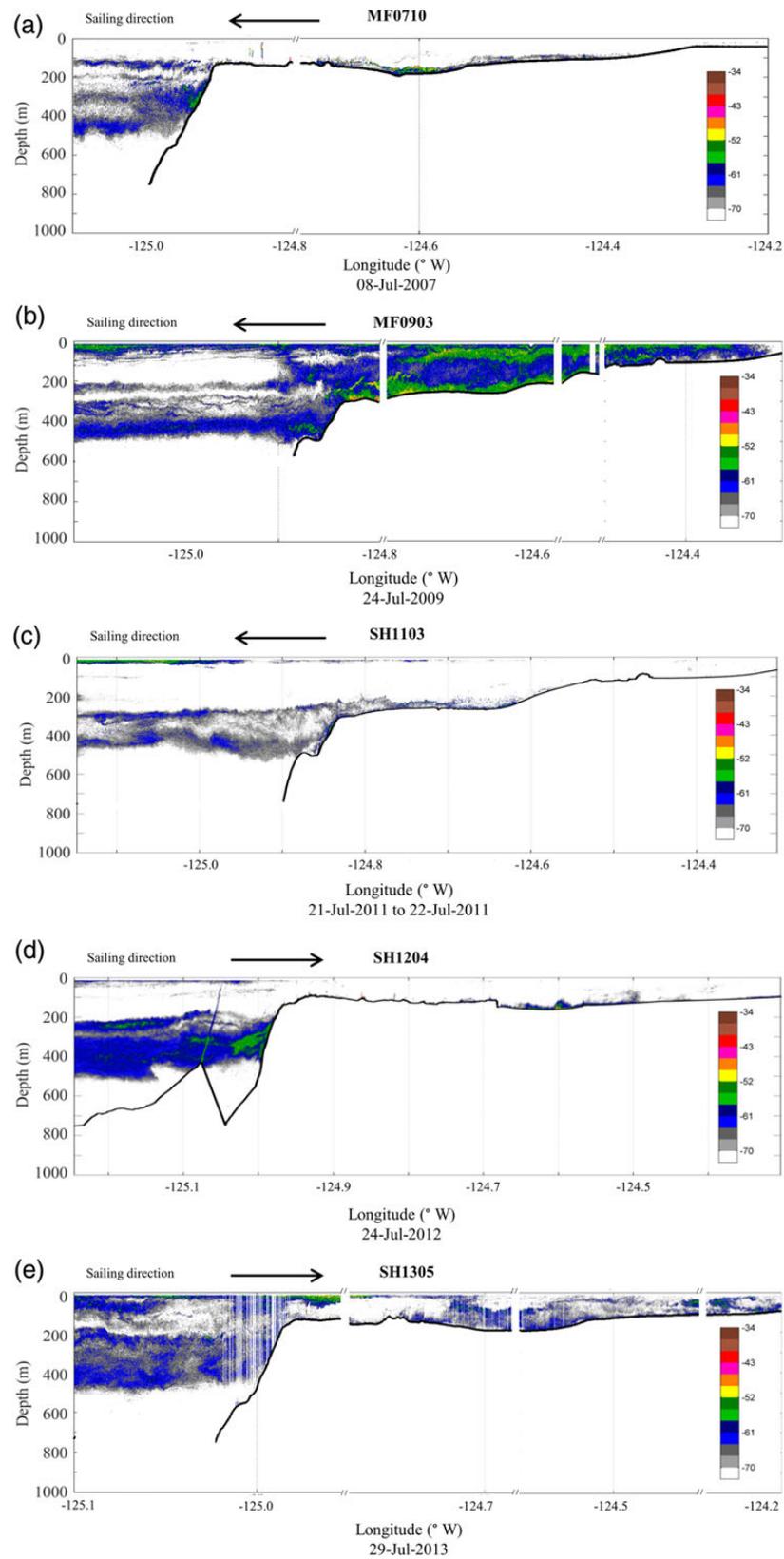
The depth of the seabed was estimated using Echoview's (Myriax Pty, Ltd, Hobart, Tasmania, Australia) best bottom candidate algorithm. Data within 1 m of and below the estimated bottom depth were removed as only the water column acoustic returns were of interest here. Processing and plotting was completed using Matlab (Mathworks, Inc., Natick, MA, USA) and data manipulation and application of the algorithm were completed using Echoview.

### Delta nautical area scattering coefficient ( $\Delta\text{NASC}$ )

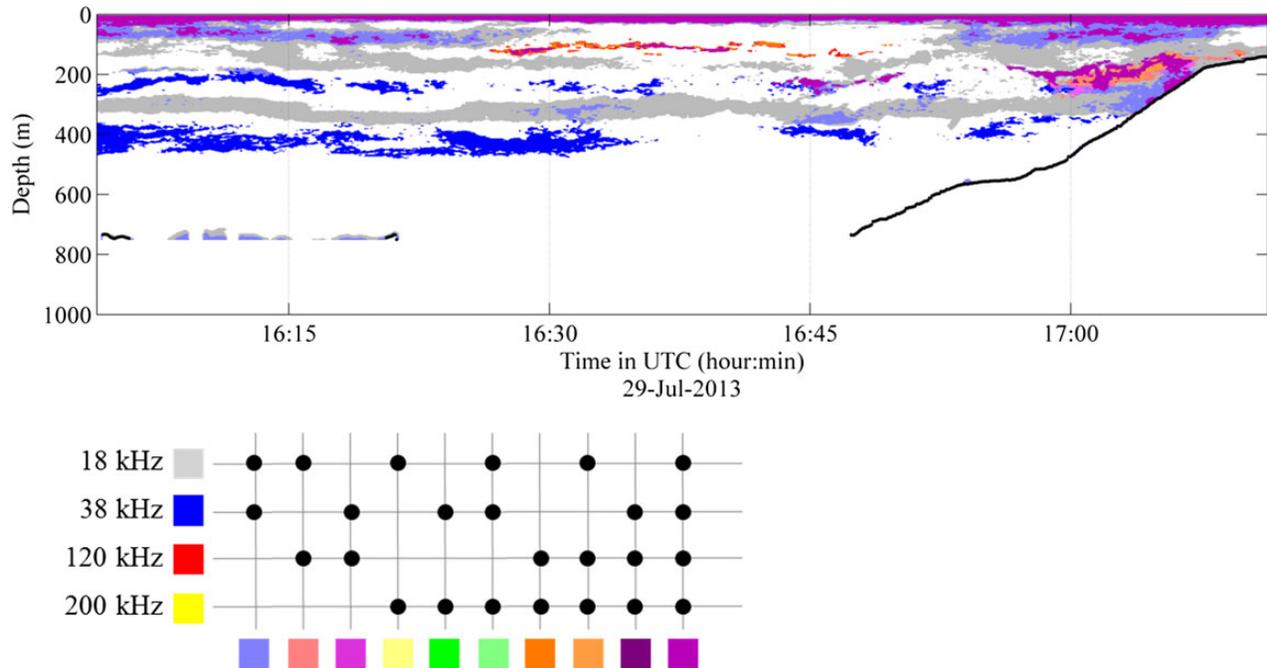
NASC ( $\text{m}^2$  nautical mile $^{-2}$ ) is an areal measure of the cumulative backscattering through a layer of water averaged over some horizontal extent (MacLennan *et al.*, 2002). For these data, we calculated NASC throughout the water column (less than the top 14 m and within 1 m of the seabed) in 250 m depth bins and at 1 nautical mile intervals. Before calculating frequency-dependent NASC, data within 1 m of and below the manually corrected seabed detection were removed to ensure strong returns from the ocean bottom did not interfere with the returned NASC values. The difference in NASC values ( $\Delta\text{NASC}$ ) between 120 and 38 kHz [ $\Delta\text{NASC} = \text{NASC}(120 \text{ kHz}) - \text{NASC}(38 \text{ kHz})$ ] was used to illustrate the relative change in reflectance properties in the water column along the transect and across years. This is analogous to the common "dB differencing" or "delta-DB" methods of separating gas-bearing fish from zooplankton species (e.g. Madureira *et al.*, 1993; Lawson *et al.*, 2008). NASC was calculated using Echoview and plotted with Matlab.

### Multifrequency indicator

The MFI algorithm is an index used for measuring species diversity. Four classifications of scatterers were extracted from the acoustic data, namely non-gas-bearing organisms, small bubbles, fluid-like



**Figure 2.**  $S_v$  data at 38 kHz along the transect of interest ( $43.872^\circ\text{N}$ ,  $125.172^\circ\text{W}$ ;  $43.975^\circ\text{N}$ ,  $124.211^\circ\text{W}$ ) for cruises conducted in (a) 2007, (b) 2009, (c) 2011, (d) 2012, and (e) 2013. Gaps in the time-series for 2007, 2009, and 2013 cruises result from the removal of off-transect data. The black line indicates the detected seabed. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.



**Figure 3.** (Upper panel) Four frequency Simrad EK60  $S_v$  collected in 2013 on the NOAA Ship Bell M. Shimada processed using the MFSBI algorithm and depicted using the legend shown in the lower panel. Black line indicates the seabed estimated using Echoview's best bottom candidate algorithm. The acoustic diversity identified in this image resulted in the selection of files on this east/west transect near the 44th parallel to be used as the region of interest (Figure 1). (Lower panel) The MFSBI colour scale applied to the imagery. The dots indicate the frequencies where  $S_v$  values were above the threshold value. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.

zooplankton, and large gas-bearing organisms (Trenkel and Berger, 2013). Non-gas-bearing organisms indicate non-swimbladder fish such as mackerel, whereas large gas-bearing organisms indicate gas-filled-swimbladder-bearing fish such as Pacific hake (*Merluccius productus*). In addition to small perturbations in the water column, small bubbles are indicative of larval fish or phytoplankton. Fluid-like zooplankton indicates euphausiids, copepods, or shrimp. Preprocessed data were converted from dB to linear scale. The MFI algorithm outlined in Trenkel and Berger (2013) was then applied. This index measures backscatter energy distribution while maintaining the order of the sonar frequencies. See Supplementary material for details of our implementation.

Classification schemes were then extracted from the resulting MFI grid based on the value ranges suggested by Figure 4 in Trenkel and Berger (2013). Large gas-bearing organisms fall within  $0.0 \leq \text{MFI} \leq 0.4$ ; small bubbles are categorized as  $0.4 < \text{MFI} \leq 0.6$ ; fluid-like zooplankton are categorized as  $0.7 \leq \text{MFI} \leq 0.8$ , and non-gas-bearing organisms range from  $0.8 < \text{MFI} \leq 1.0$ . The parameters applied were optimized for these data by visually scrutinizing a selection of echograms. The MFI algorithm was applied using Echoview and the results were plotted using Matlab. The MFI classifications are displayed as boxplots of the  $S_v$  values for each classification where the median, 25th, 75th, and 99th percentiles and outliers are presented to highlight the spatial patterns of the acoustic classification with respect to the continental shelf break and annual changes within and across acoustic categories.

## Results

### Multifrequency single-beam imagery (MFSBI)

The MFSBI algorithm applied to the subset of  $S_v$  data within the region of interest along the 44th parallel illustrates a diversity of content

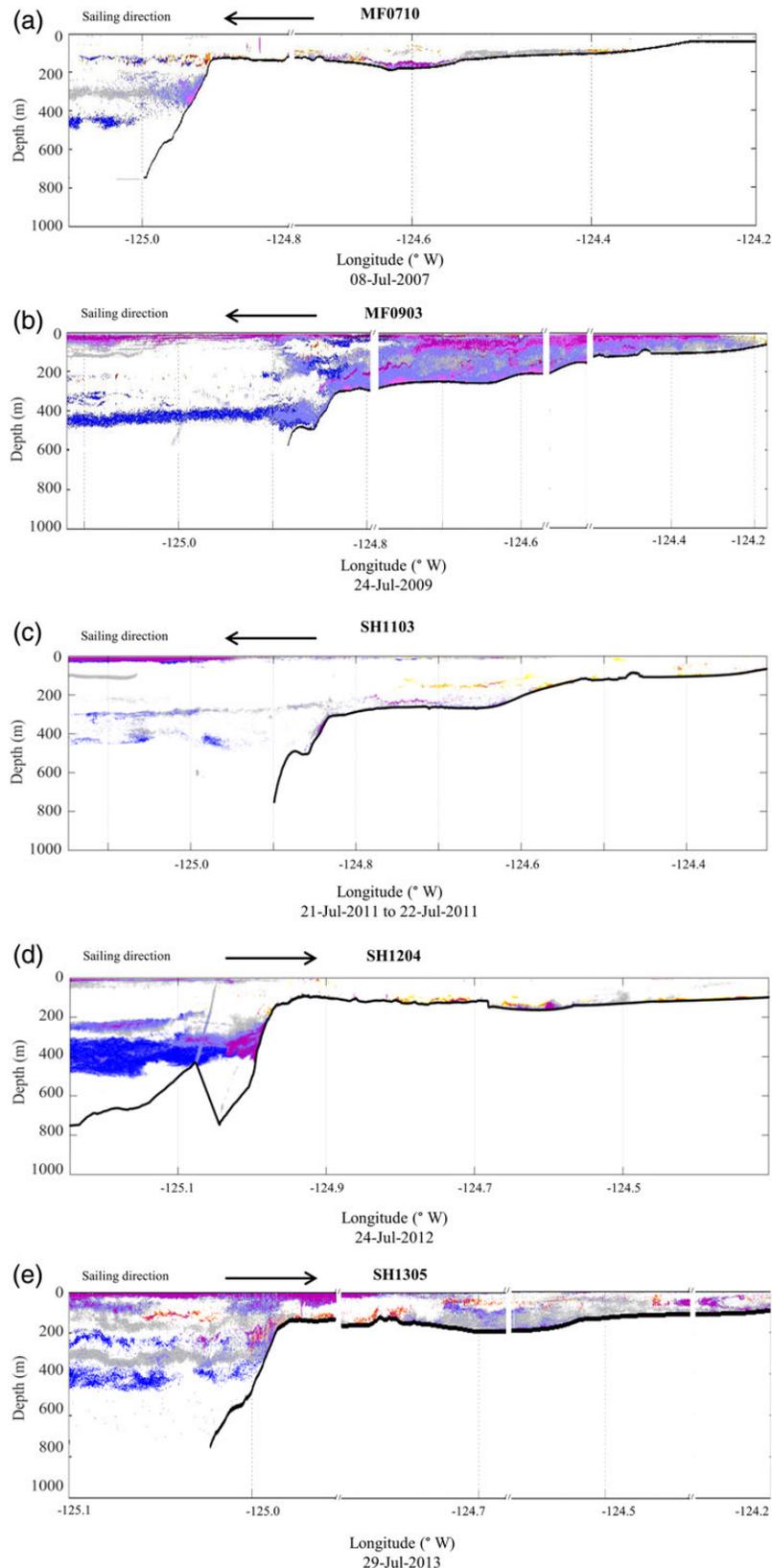
in the water column and variability among years (Figure 4). Despite interannual variation, elevated backscatter at and beyond the continental shelf break were consistently observed for most years. Assemblages of gas-filled swimbladder-bearing fish (depicted in the low frequencies, blue/grey pixels) and, to a lesser degree, layers of zooplankton (depicted in the high frequencies, reds/orange pixels) extend off-the-shelf mainly between 400 and 600 m water depth with some layers also present near the surface. The 2009 transect illustrates the greatest level of backscatter along the shelf compared with the other years.

### Delta nautical area scattering coefficient ( $\Delta\text{NASC}$ )

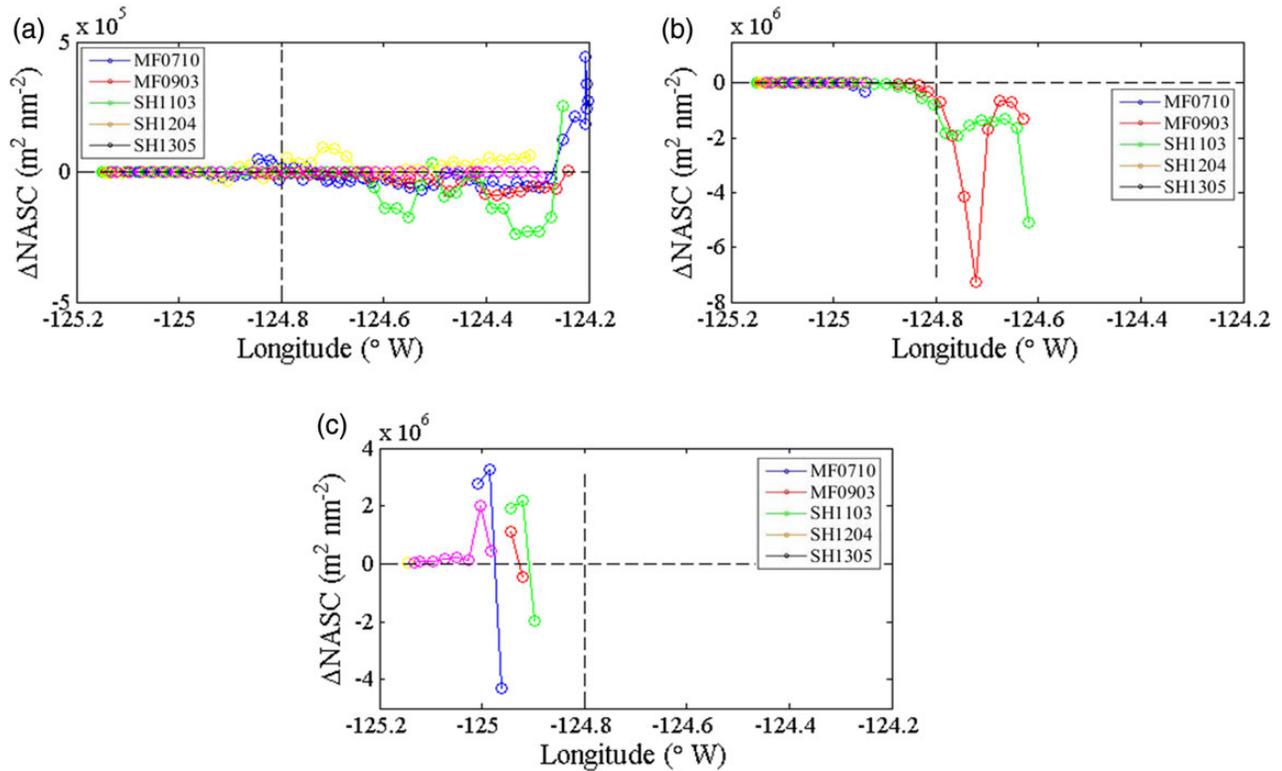
The  $\Delta\text{NASC}$  analysis depicts the difference between the 120 and 38 kHz throughout the water column. Most data points fall below zero indicating a stronger backscatter within 38 kHz (greater NASC value compared with 120 kHz). However, near surface (0–250 m depth bin)  $\Delta\text{NASC}$  values near the coast fall above zero, suggesting a greater proportion of smaller fluid-like scatterers (e.g. euphausiids, shrimp; Figure 5a). The  $\Delta\text{NASC}$  values integrated between 250–500 m water depth show large negative spikes, suggesting a dominance of juvenile and/or adult fish (with and without swimbladders) and larger fluid-like scatterers (e.g. squid), especially for MF0903 and SH1103 (Figure 5b).  $\Delta\text{NASC}$  at depth (500–750 m) vary greatly with a shift towards increased 120 kHz and/or decreased 38 kHz backscatter (Figure 5c). Increased backscatter at 120 kHz within this depth range is a result of noise not fully removed by the background and impulsive filters.

### Multifrequency indicator

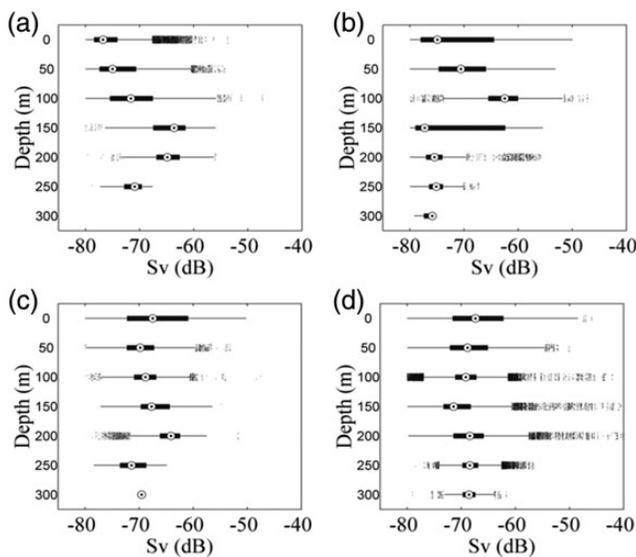
Despite the application of filters to remove background noise and spurious signals to reduce acoustic noise, the most prevalent feature observed in the MFI algorithm analysis was a persistent



**Figure 4.** Application of the MFSBI algorithm to the subset of data extracted along the transect of interest ( $43.872^{\circ}\text{N}$ ,  $125.172^{\circ}\text{W}$ ;  $43.975^{\circ}\text{N}$ ,  $124.211^{\circ}\text{W}$ ) for cruises conducted in (a) 2007, (b) 2009, (c) 2011, (d) 2012, and (e) 2013. The legend for these images is presented in the bottom panel of Figure 3. Gaps in the time-series for 2007, 2009, and 2013 cruises result from the removal of off-transect data. The black line indicates the detected seabed. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.



**Figure 5.**  $\Delta$ NASC per nautical mile for the NWFSC acoustic cruises within the region of interest. Values reflect the difference between the NASC calculated for 120 and 38 kHz between (a) 0 and 250 m depth, (b) 250 and 500 m depth, and (c) 500 and 750 m depth. Data points above zero indicate a higher NASC value at 120 kHz, while points below zero indicate a higher NASC at 38 kHz for that nautical mile. Note the difference in scale. Horizontal dashed line delineates 0  $\Delta$ NASC. Vertical dashed line indicates the approximate location of the shelf edge,  $\sim$ 250 m depth. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.



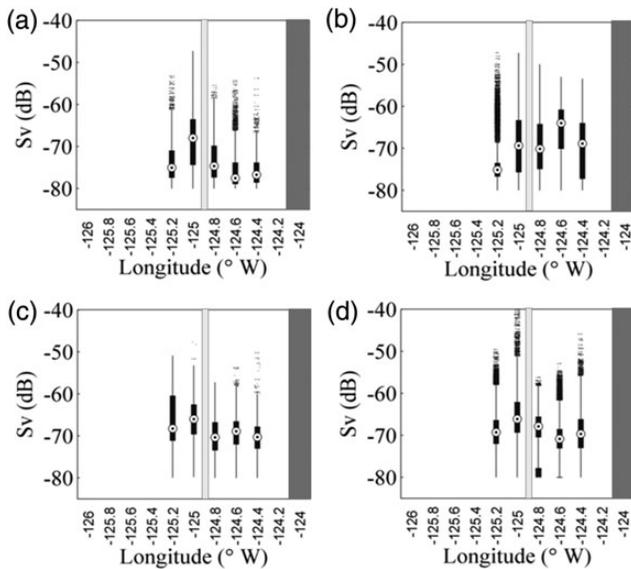
**Figure 6.** MFI results for (a) non-gas-bearing, (b) fluid-like zooplankton, (c) small bubbles, and (d) large gas-bearing calculated from SH1305 using four frequencies (18, 38, 120, and 200 kHz) binned vertically to 300 m depth. Shown are the median (black dot), 25th and 75th percentiles (black bar), 99th percentile (black line), and outliers (black dashes).

artefact at depth for all years and classifications. Even with a restrictive background noise reduction algorithm, the indices are heavily influenced by noise deeper than  $\sim$ 300–350 m.  $S_v$  at all depths was calculated for all years (SM Figures 1–4); however, only the results of this analysis for SH1305 data limited to depths above 300 m are presented (Figures 6 and 7).

Non-gas-bearing organisms (MFI between 0.8 and 1.0) resolved above 300 m for SH1305 indicate the strongest presence between 150 and 200 m depth just westward of the shelf break. Fluid-like zooplankton layers (MFI between 0.7 and 0.8) are present across the shelf with the highest backscatter volume around 100 m depth. The small bubbles classification (MFI between 0.4 and 0.6) maintained relatively low median  $S_v$  values from the surface to the shelf break across the transect. Large gas-bearing organisms (MFI between 0.0 and 0.4) showed the greatest extent of outliers by depth and by longitude. This greater range of  $S_v$  values and trend towards higher acoustic backscattering suggests a wider and stronger prevalence of this classification throughout the water column.

### Discussion

Several concepts to illustrate the quality and composition of water column sonar data were explored using data collected off the US northwest coast by the NMFS Northwest Fisheries Science Center. An adaptation of the [Jech and Michaels \(2006\)](#) algorithm with a novel colour scale (MFSBI) highlighted the fine-scale (i.e. meter



**Figure 7.** MFI results for (a) non-gas-bearing, (b) fluid-like zooplankton, (c) small bubbles, and (d) large gas-bearing calculated from SH1305 using four frequencies (18, 38, 120, and 200 kHz) binned horizontally. Shown are the median (black dot), 25th and 75th percentiles (black bar), 99th percentile (black line), and outliers (black dashes). Approximate locations of the coastline (dark grey bar) and continental shelf break (light grey bar) are also shown.

vertically by ping horizontally) composition of multifrequency single-beam data in a single image with greatest adaptability for automated processing. Here, the dominating frequency or frequencies are depicted per pixel providing an understanding of changes in content with time and space (depth and location).

The  $\Delta$ NASC values integrate the water column and thus illustrate large-scale areal (i.e. latitudinal and longitudinal) patterns of changes in the two frequencies (38 and 120 kHz) that indicate distinction between fluid-like zooplankton (e.g. krill) and gas-bearing organisms such as fish. This metric is intended to visualize large-scale areal trends, which can be discerned from this analysis; for example, increased variability at and beyond the continental shelf break. As such, interpretation of fine-scale patterns is more difficult than with the MFSBI or MFI algorithms.

Implementation of the MFI algorithm and its interpretation were taken directly from [Trenkel and Berger \(2013\)](#) and one assumption is that their classification schemes are directly transferable to marine communities in other parts of the world. More research is needed to verify this, but there is theoretical basis for at least a first-order approximation that the acoustic backscatter among taxa or scattering types (i.e. fluid-like, gas bearing) is similar and common among the world's oceans (i.e. [Lavery et al., 2007](#); [Lawson et al., 2008](#)). An advantage of the MFI to the MFSBI method is that it is an empirically based (and loosely theoretically based) statistical approach to classification.

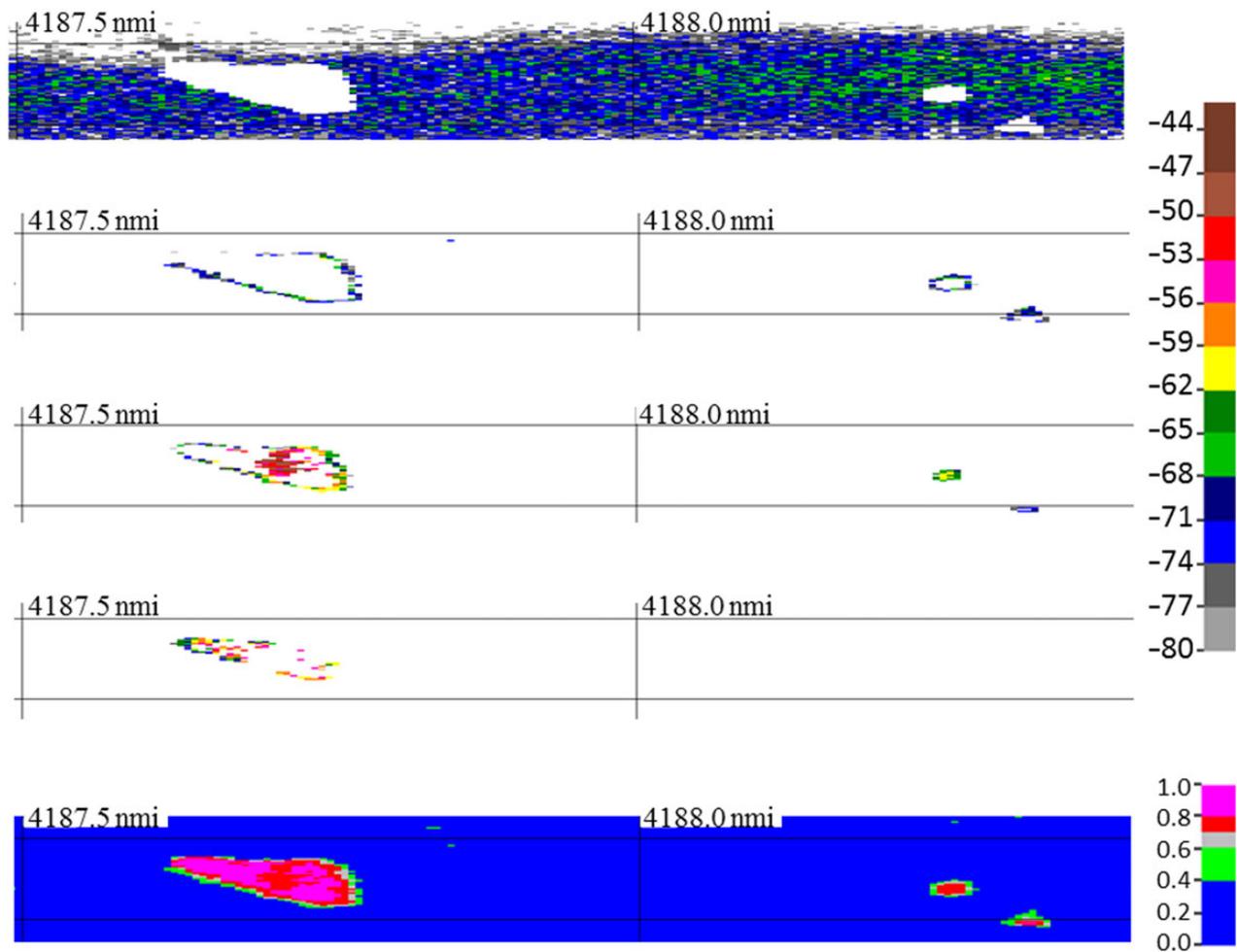
Application of the MFI algorithm differentiates four classes of acoustic scatterers to provide a depiction of the content of the water column. For example, elevated median  $S_v$  values in the large gas-bearing classification near the shelf break and between 100 and 200 m depth in 2013 are supported by the presence of backscatter for low frequencies (blue/grey pixel) observed off-the-shelf in the 2013 MFSBI imagery. However, extensive noise below 300 m

prevented an accurate understanding of the classifications below this depth for all years, and future developments would include automated selection of frequencies based on noise characteristics. For example, the 200 kHz data become more noise limited at greater depths than the other frequencies, so a depth-dependent MFI (or other indicator) could be implemented where higher frequencies are removed with increasing depth and potentially only one to two frequencies are used at full ocean depth. Accurate delineation of each acoustic classification was further blurred by the presence of “halos” around the zooplankton schools as observed through visual scrutiny (Figure 8). This effect is suspected to be due to differences in the beam pattern between the 18-kHz transducer ( $11^\circ$ ) and the other transducers ( $7^\circ$ ; *sensu* [Diner, 2001](#)). Remnants of zooplankton schools across classifications could have a substantial effect on results. Further work is needed to refine the filtering and MFI parameters to eliminate the impact of these sources of error.

The NWFSC conducts an integrated acoustic and trawl (IAT) survey to assess the distribution, abundance, and biology of coastal Pacific hake along the Pacific coast of the United States and Canada. This survey has been conducted by the NWFSC since 2001 ([Fleischer et al., 2005](#)) and the Alaska Fisheries Science Center since 1997 ([Wilson et al., 2000](#)). This survey is intended to correspond to the seasonal migration from offshore and south, where hake spend the winter spawning season, to the north and coastal areas between northern California and northern British Columbia during spring, summer, and fall. The success of the survey is predicated in part by the coincident timing with the migration, and changes in migration behaviour and/or spatio-temporal patterns could influence the accuracy of the estimates from the survey. While we have not attempted to undertake a full analysis of the survey, acoustic backscattering patterns observed using the three methods presented here suggest a consistency with historical patterns. For example, [Swartzman \(2001\)](#) observed an overlap of actively feeding hake and large patches of euphausiids, their dominant prey, at or near the shelf break (150–250 m depth). This pattern is consistently observed using the methods presented here, and provides the potential for a cohesive analysis of historical data.

The overall goal of these efforts is to establish effective methods to analyse and visualize over 21 TB of archived water column sonar data for public dissemination and improved understanding of the marine communities and their habitat. Over 40% of the total data archived to date derives from Simrad EK60 sonar systems and the concepts presented here are candidates for wide-spread application and integration into the archive's data access web page. In addition to ease of understanding, additional aspects to consider include the algorithm's flexibility, robustness, adaptability across regions and datasets, and potential for batch processing.

Accurately detecting the seabed using automated algorithms remains a persistent problem due to demersal assemblages and presence of strong acoustic noise in the water column, both of which can obfuscate seabed detection. An improperly identified seabed can have profound effects on the results of the NASC and MFI analyses, and thus requires manual seabed scrutinization or at least tuning to remove erroneous data points. This process can be one of the most time-intensive components of the post-processing, and advances in accurate seabed detection will be a great time savings to the improvement and efficiency of automated processing and analysis of water column sonar data. The imagery created using MFSBI algorithm is conducive to automated seabed detection, though inaccuracies due to noise would result in the presence of spurious data or the exclusion of potentially legitimate water column data (see Figure 4d



**Figure 8.** Example of the “halo” effect from the MFI algorithm identified in an MF0903 file. (a–d) “Zoomed-in”  $S_v$  echograms at 120 kHz highlighting suspected zooplankton patches within the four MFI classifications: (a) large gas-bearing organisms, (b) small bubbles, (c) fluid-like zooplankton, and (d) non-gas-bearing organisms, and (e) is the MFI grid. This figure is available in black and white in print and in colour at *ICES Journal of Marine Science* online.

at approximately  $-125.5^\circ\text{W}$ ). Attempts have been made to optimize the seabed detection parameters; however, without *a priori* knowledge of the extent of each cruise’s depth range and data quality, it is impossible to account for every condition that could arise in the archived sonar data, which are collected across the nation and even globally. Integration of high-resolution bathymetric data available through NCEI to create a bottom mask for surveys covering known areas will be explored. The mask would remove data below the bathymetrically defined seabed and serve as a first step before or in place of manual detection. It could also help inform the automated seabed detection algorithm for improved accuracy. Issues with comparative resolution (less than a metre for the acoustic data and tens of metres to a kilometre for the bathymetry data) will have to be considered, especially in areas with steep gradients such as the shelf break.

Even with accurate seabed detection, the quality of the raw acoustic data is pivotal in determining the quality and benefit of the resulting processed imagery. Noise or attenuation can greatly affect the output and is often the cause for the seabed detection algorithm to fail. The preprocessing methodology implemented here only accounted for impulsive and background noise. Transient noise and signal attenuation are also prevalent sources of error that

should be addressed. Similar to seabed detection, robust and automated or semi-automated algorithms are strongly needed to efficiently filter such noise. When done effectively, the quality of the data can be greatly improved (Ryan *et al.*, 2015). Transducer resonance or ring-down is another significant source of noise, especially in the lower frequencies, that could be a major component of the total noise in environments with changing bottom (e.g. continental shelves and seamounts). Ryan *et al.* (2015) demonstrate the application of impulsive, transient, and attenuation filters to basin-scale data in the Southern Ocean. Future implementation of algorithms that automatically remove such noise should improve the quality of the processed data.

Manual examination is needed to properly delineate the acoustic classifications for all methods when applied to different regions. For these reasons, the size of the subset of data analysed was greatly limited to a manageable number of files in anticipation of human scrutiny. Due to the volume of data to process the whole archive, such manual tuning quickly becomes overwhelming and impractical. Thus, it is imperative that statistically robust algorithms based on empirical data and theoretical models be developed so that automated processing becomes less time intensive and more effort can be afforded to developing interpretation and analysis

methods, which can be applied to ecological and management issues and questions.

All methods have benefits and limitations. The MFSBI algorithm proved most suitable for batch processing and broad application for initial visualization of the data. This method transforms complex data stored in proprietary file formats into a digestible image or graphic and is highly valuable for a broad audience of varying backgrounds. Even the most novice users will be able to quickly identify what datasets are of potential interest and relevant to answering their research objectives before requesting large volumes of data. One of the limitations of the MFSBI algorithm is the lack of statistical robustness for classification. The MFI algorithm will require more effort in the near term to evaluate and assess its accuracy and precision (i.e. taxonomic level) of classification, but it is a step towards a statistically robust method and may prove useful for comparing marine communities on a global scale. The  $\Delta$ NASC algorithm masks vertical patterns but is valuable for presenting large-scale (i.e. regional to basin-wide scales) features and patterns.

### Future efforts

The ability to process large volumes of data to address issues and questions at the scales that may be required to manage living marine resources have been limited by access to data and methods to efficiently process those data. Access to these data is now not an issue. As more data are archived from the different parts of the United States and as more countries archive their data, impediments to accessing large volumes will continue to decrease. The next step will be to develop methods to process these data. We have only scratched the surface here using three existing methods; there are many more that have been or will be developed. Additional methods to process large volumes of data for acoustic features will be investigated for their robustness and potential to be applied to the archived data. Two recently published methods for automatically identifying sound-scattering layers (SSL) could be used to understand basin-scale SSL characteristics over multiple years (Cade and Benoit-Bird, 2014; Proud *et al.*, 2015). Efforts to improve the linkage to concurrently collected oceanographic (physical and biological) and bathymetric data are underway to provide users with an ecosystem-wide understanding of the area ensonified. These are just two examples of many that can be explored to increase the value of the archived data. With access to large datasets spanning wide temporal and spatial scales and through improved processing, researchers will be able to address new scientific questions in a cost-effective and efficient manner.

### Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

### Acknowledgements

This project was funded by the NOAA National Marine Fisheries Service and NOAA National Environmental Satellite, Data, and Information Service. The authors thank Charles Anderson for providing the data from the NOAA NCEI water column sonar data archive. Any use of trade names does not imply endorsement by NOAA.

### References

Andersen, L. N. 2001. The new Simrad EK60 scientific echo sounder system. *The Journal of the Acoustical Society of America*, 109: 2336.

- Benoit-Bird, K. J., and Au, W. W. 2001. Target strength measurements of Hawaiian mesopelagic boundary community animals. *The Journal of the Acoustical Society of America*, 110: 812–819.
- Cade, D. E., and Benoit-Bird, K. J. 2014. An automatic and quantitative approach to the detection and tracking of acoustic scattering layers. *Limnology and Oceanography: Methods*, 12: 742–756.
- Cutter, G. R., Berger, L., and Demer, D. A. 2010. A comparison of bathymetry mapped with the Simrad ME70 multibeam echosounder operated in bathymetric and fisheries modes. *ICES Journal of Marine Science*, 67: 1301–1309.
- De Robertis, A., and Higginbottom, I. 2007. A post-processing technique to estimate the signal-to-noise ratio and remove echosounder background noise. *ICES Journal of Marine Science*, 64: 1282–1291.
- Diner, N. 2001. Correction on school geometry and density: approach based on acoustic image simulation. *Aquatic Living Resources*, 14: 211–222.
- Fleischer, G. W., Cooke, K., Ressler, P. H., Thomas, R. E., de Blois, S. K., Hufnagle, L. C., Kronlund, A. R., *et al.* 2005. The 2003 integrated acoustic and trawl survey of Pacific hake, *Merluccius productus*, in U.S. and Canadian waters off the Pacific coast. US Department of Commerce, NOAA Technical Memorandum, NMFS-NWFSC-65.
- Foote, K. G. 2009. Acoustic methods: brief review and prospects for advancing fisheries research. *In The Future of Fisheries Science in North America*, pp. 313–343. Ed. by R. J. Beamish, and B. J. Rothschild. Springer, The Netherlands.
- Foote, K. G., Knudsen, H., Vestnes, G., MacLennan, D. N., and Simmonds, E. J. 1987. Calibration of acoustic instruments for fish density estimation: a practical guide. Cooperative Research Report, 144. ICES, Palaegade, Copenhagen, Denmark.
- Hutin, E., Simard, Y., and Archambault, P. 2005. Acoustic detection of a scallop bed from a single-beam echosounder in the St Lawrence. *ICES Journal of Marine Science*, 62: 966–983.
- ICES. 2014. A metadata convention for processed acoustic data from active acoustic systems. SISP 4 TG-AcMeta, ICES WGFASST Topic Group, TG-ACMeta.
- Jech, J. M., and Michaels, W. L. 2006. A multifrequency method to classify and evaluate fisheries acoustics data. *Canadian Journal of Fisheries and Aquatic Sciences*, 63: 2225–2235.
- Lavery, A. C., Chu, D., and Moum, J. N. 2010. Measurements of acoustic scattering from zooplankton and oceanic microstructure using a broadband echosounder. *ICES Journal of Marine Science*, 67: 379–394.
- Lavery, A. C., Wiebe, P. H., Stanton, T. K., Lawson, G. L., Benfield, M. C., and Copley, N. 2007. Determining dominant scatterers of sound in mixed zooplankton populations. *The Journal of the Acoustical Society of America*, 122: 3304–3326.
- Lawson, G. L., Wiebe, P. H., Stanton, T. K., and Ashjian, C. J. 2008. Euphausiid distribution along the Western Antarctic Peninsula—Part A: development of robust multi-frequency acoustic techniques to identify euphausiid aggregations and quantify euphausiid size, abundance, and biomass. *Deep Sea Research II: Topical Studies in Oceanography*, 55: 412–431.
- MacLennan, D. N., Fernandes, P. G., and Dalen, J. 2002. A consistent approach to definitions and symbols in fisheries acoustics. *ICES Journal of Marine Science*, 59: 365–369.
- Madureira, L. S., Everson, I., and Murphy, E. J. 1993. Interpretation of acoustic data at two frequencies to discriminate between Antarctic krill (*Euphausia superba* Dana) and other scatterers. *Journal of Plankton Research*, 15: 787–802.
- Misund, O. 1997. Underwater acoustics in marine fisheries and fisheries research. Review in *Fish Biology and Fishes*, 7: 1–34.
- NWFSC. 2007. Water Column Sonar Data Collection (MF0710, EK60). National Centers for Environmental Information, NOAA. doi:10.7289/V5Q81B1B. Accessed 1 May 2015.
- NWFSC. 2009. Water Column Sonar Data Collection (MF0903, EK60). National Centers for Environmental Information, NOAA. doi:10.7289/V5KK98QC. Accessed 1 May 2015.

- NWFSC. 2011. Water Column Sonar Data Collection (SH1103, EK60). National Centers for Environmental Information, NOAA. doi:10.7289/V53J39XH. Accessed 1 May 2015.
- NWFSC. 2012. Water Column Sonar Data Collection (SH1204, EK60). National Centers for Environmental Information, NOAA. doi:10.7289/V5ZS2TF7. Accessed 1 May 2015.
- NWFSC. 2013. Water Column Sonar Data Collection (SH1305, EK60). National Centers for Environmental Information, NOAA. doi:10.7289/V57942N7. Accessed 22 April 2015.
- Pirtle, J. L., Weber, T. C., Wilson, C. D., and Rooper, C. N. 2015. Assessment of trawlable and untrawlable seafloor using multibeam-derived metrics. *Methods in Oceanography*, 12: 18–35.
- Proud, R., Cox, M. J., Wotherspoon, S., and Brierley, A. S. 2015. A method for identifying sound scattering layers and extracting key characteristics. *Methods in Ecology and Evolution*, 6: 1190–1198.
- Ryan, T. E., Downie, R. A., Kloser, R. J., and Keith, G. 2015. Reducing bias due to noise and attenuation in open-ocean echo integration data. *ICES Journal of Marine Science*, 72: 2482–2493.
- Simmonds, E., and MacLennan, D. 2005. *Fisheries Acoustics: Theory and Practice*, 2nd edn. Fish and Fisheries. Blackwell Publishing, Oxford. 456 pp.
- Skarke, A., Ruppel, C., Kodis, M., Brothers, D., and Lobecker, E. 2014. Widespread methane leakage from the sea floor on the northern US Atlantic margin. *Nature Geoscience*, 7: 657–661.
- Soria, M., Fréon, P., and Gerlotto, F. 1996. Analysis of vessel influence on spatial behaviour of fish schools using a multi-beam sonar and consequences for biomass estimates by echo-sounder. *ICES Journal of Marine Science*, 53: 453–458.
- Stanton, T. K., Chu, D., Jech, J. M., and Irish, J. D. 2010. New broadband methods for resonance classification and high-resolution imagery of fish with swimbladders using a modified commercial broadband echosounder. *ICES Journal of Marine Science*, 67: 365–378.
- Swartzman, G. 2001. Spatial patterns of Pacific hake (*Merluccius productus*) shoals and euphausiid patches in the California Current Ecosystem. In *Spatial Processes and Management of Marine Populations*, pp. 495–512. Ed. by G. H. Kruse, N. Bez, A. Booth, M. W. Dorn, S. Hills, R. N. Lipcius, D. Pelletier, *et al.* University of Alaska Sea Grant, AK-SG-01-02, Fairbanks.
- Trenkel, V. M., and Berger, L. 2013. A fisheries acoustic multi-frequency indicator to inform on large scale spatial patterns of aquatic pelagic ecosystems. *Ecological Indicators*, 30: 72–79.
- Weber, T. C., Jerram, K., and Mayer, L. 2014. Acoustic sensing of gas seeps in the deep ocean with split-beam echosounders. In *Proceedings of Meetings on Acoustics*, p. 070057. Acoustical Society of America.
- Weber, T. C., Mayer, L., Beaudoin, J., Jerram, K., Malik, M., Shedd, B., and Rice, G. 2012. Mapping gas seeps with the deepwater multibeam echosounder on Okeanos Explorer. *Oceanography*, 25: 54–55.
- Wilson, C. D., Guttormsen, M. A., Cooke, K., Saunders, M. W., and Kieser, R. 2000. Echo integration-trawl survey of Pacific hake, *Merluccius productus*, off the Pacific coast of the United States and Canada during July–August 1998. US Department of Commerce, NOAA Technical Memorandum, NMFS-AFSC-118.

Handling editor: Richard O'Driscoll